The Annals of Applied Statistics
2011, Vol. 5, No. 1, 91–95
DOI: 10.1214/10-AOAS398J
Main article DOI: 10.1214/10-AOAS398

DISCUSSION OF: A STATISTICAL ANALYSIS OF MULTIPLE TEMPERATURE PROXIES: ARE RECONSTRUCTIONS OF SURFACE TEMPERATURES OVER THE LAST 1000 YEARS RELIABLE?[1]

By Eugene R. Wahl and Caspar M. Ammann

NOAA's National Climatic Data Center and the National Center for Atmospheric Research

McShane and Wyner (2011) (MW) introduce new methods into the effort to reconstruct the large-scale Northern Hemisphere temperature average over the past millennium, helping to advance interaction between applied statisticians and paleoclimatologists to improve understanding of preinstrumental climates. However, despite a good effort to capture the various points of contention in the reconstruction arena, MW provide an incomplete, and at times inadequate, review of the existing literature considering reconstruction of Northern Hemisphere surface temperatures over the past millennium. In particular, the evaluations cited regarding the original Mann, Bradley and Hughes (MBH) reconstruction (1998/1999) and MBH's use of principal component (PC) summaries of dendroclimatic proxy data fail to address this issue properly, and in the process propagate errors that have been fully addressed in the literature. Similarly, MW omit important information in their examination of the methodology outlined and used by Ammann and Wahl (2007) (AW) to test the significance of the MBH reconstruction. Because examinations related to the MBH reconstruction have had particular salience not only in the specialist literature, but also politically [cf. House Committee on Energy and Commerce (2005); American Association for the Advancement of Science (2005); Wegman, Scott and Said (2006); Russell et al. (2010)] and in terms of scientific review and advisement [National Research Council (2006)], it is important that these omissions be corrected. We will focus our discussion on them, along with providing more general closing observations.

Received September 2010; revised September 2010.

[1]We gratefully acknowledge the support of the Weather and Climate Impacts Assessment Science Program at the National Center for Atmospheric Research. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

MBH used PC summaries of tree ring proxy information in different parts of the world to reduce the weighting of these regions in terms of the number of proxy data series from them employed in the MBH reconstruction. The use of PCs (more generally, the use of eigenvector/singular value decomposition methods) for the purpose of dimension reduction of data is a common practice in climatology/paleoclimatology. The bulk of attention about the use of PC summarization by MBH has focused on a network of 70 tree ring data series from North America used in their reconstruction in the first half of the 15th century (called ITRDB, from the International Tree Ring Data Base), and is associated with significant confusion and erroneous claims about the extent to which the way MBH employed this practice biased their results. MW add to this confusion.

First, it should be noted that MBH formed the underlying proxy data into standardized anomalies relative to their calibration period, 1902–1980, rather than centering and standardizing over the full length of the data. This was done because the MBH reconstruction consists of 12 segments, each of which was recalibrated to include increasing numbers of proxy data series as time comes closer to the present. Their practice ensured that all the proxy data *and* the instrumental data used in calibration were processed in common terms. This was a reasonable judgment given the segmental nature of the MBH reconstruction, although the normal methodology for PC extraction from standardized anomaly data is to center and standardize over the full data period.

In AW (Supplement, Section 2; http://www.cgd.ucar.edu/ccr/ammann/millennium/AW_supplement.html) we evaluate the impact these two methods of data treatment have on the extraction of PCs. In the MBH "common centered" method, the first PC contains a noticeable "hockey stick"-like shape, whereas in the "full centered" method this shape is spread across the first two PCs. When a vector sum of the first two PCs from both methods is calculated, the plots of the resultant time series have essentially identical shapes, with a larger amplitude in the case of the "common centered" method. A point that is generally not appreciated in this context [e.g., National Research Council (2006); MW], is that MBH used the first *two* North American ITRDB PCs in their reconstruction of the time period 1400–1449, which is the key target of critical focus in this regard [cf. McIntyre and McKitrick (2003, 2005a, 2005b, 2005c)] (MM). In Wahl and Ammann (2007) (WA), we systematically examined the difference the use of the first two PCs from each method actually has in the MBH reconstruction for this time period, and note that the reconstructed time series' structures are nearly identical, with a slight average warming of 0.05 deg. C when "common centered" PCs are used (WA, Figure 3, blue range; cf. page 51).

When nonstandardized anomalies are input into a PC algorithm using the variance–covariance matrix rather than the correlation matrix (as MM;

cf. WA, page 45), the "hockey-stick"-like shape occurs lower in the PC order, here PC4. This occurs because the algorithm captures information in the first one or two eigenvectors/PCs that is primarily related to the absolute magnitude of the largest-scaled variables in the data, which here differ by a factor of 13 (AW, Supplement, Sections 2, 4). In WA, we show that the impact of using only the first two or three ITRDB PCs extracted this way is a different reconstruction (pink curve, Figure 3), which does not pass validation testing. When the fourth or fifth ITRDB PCs are added, the reconstruction converges to those using two PCs derived from fully standardized data, all of which pass validation (WA, Figure 3, blue range; cf. AW, Supplement, Section 1).

Concerning the significance testing method outlined and used by AW, we would like to re-emphasize the consideration that using the full AR structure of proxy data to drive creation of "pseudo-proxies" as null models for establishing reconstruction skill significance thresholds is likely overly conservative. More importantly, MW omit from their examination of our reasoning that we report application of this (in our view) overly conservative method in pseudo-proxy generation (AW, pages 77–78). Thus, the kind of test MW emphasize, when applied with real-world context to our emulation of the MBH reconstruction, shows highly successful validation results. Ten of the 12 segments are significant at the 95% level (the other two are significant at the 89% and 94% levels), and the much-discussed 1400–1449 and 1450–1499 segments are significant at 99% and 96%, respectively. We wish to emphasize this point, as it distinctly counters MW's argument concerning this issue.

Finally, there is an extensive literature contradicting MW's assertions about low or poor relationships between proxies and climate. The climate system certainly exhibits spatio-temporal complexity. Yet global/hemispheric average temperature is far from being "nonpredictable." With indications about the evolution of a few primary modes of variability and independent knowledge of key radiative forcing histories, one can explain the mean temperature evolution, including its spatial expression, over the time period of instrumental data [Tett et al. (1999); Ammann et al. (2003); Hansen et al. (2007)]. Such relationships should hold as well for the last millennium, which is well illustrated in Figure 6.14 of Jansen et al. (2007); note that the figure demonstrates this result is robust to uncertainty concerning the strength of the solar irradiance forcing data. The challenge in reconstructions remains the proper implementation of the available geophysical knowledge (cf. AW), including the cross-dependence of its uncertainties. When random samples systematically outperform predictors known to contain useful information, such a result suggests the need to reevaluate the reconstruction model itself (e.g., LASSO). Apparently, important aspects were missed or improperly implemented. Otherwise, independent proxy evaluations would long ago have drawn the general conclusion that there is no skill in capturing spatial and temporal climate, which is not borne out by the literature.

REFERENCES

American Association for the Advancement of Science (2005). Letter from Alan I. Lesher, Chief Executive Officer, to Honorable Joe Barton, Chair, House Committee on Energy and Commerce, July 13.

Ammann, C. M. and Wahl, E. R. (2007). The importance of the geophysical context in statistical evaluations of climate reconstruction procedures. *Climatic Change* **85** 71–88.

Ammann, C. M., Meehl, G. A., Washington, W. M. and Zender, C. S. (2003). A monthly and latitudinally varying volcanic forcing dataset in simulations of 20th century climate. *Geophysical Research Letters* **30** 1657.

Hansen, J., Sato, M., Ruedy, R., Kharecha, P., Lacis, A., Miller, R., Nazarenko, L., Lo, K., Schmidt, G. A., Russell, G., Aleinov, I., Bauer, S., Baum, E., Cairns, B., Canuto, V., Chandler, M., Cheng, Y., Cohen, A., Del Genio, A., Faluvegi, G., Fleming, E., Friend, A., Hall, T., Jackman, C., Jonas, J., Kelley, M., Kiang, N. Y., Koch, D., Labow, G., Lerner, J., Menon, S., Novakov, T., Oinas, V., Perlwitz, J., Perlwitz, J., Rind, D., Romanou, A., Schmunk, R., Shindell, D., Stone, P., Sun, S., Streets, D., Tausnev, N., Thresher, D., Unger, N., Yao, M. and Zhang, S. (2007). Climate simulations for 1880–2003 with GISS modelE. *Climate Dynamics* **29** 661–696.

House Committee on Energy and Commerce (2005). Letter to Dr. Rajendra Pachauri, Chair, Intergovernmental Panel on Climate Change, June 23. Available at http://republicans.energycommerce.house.gov/108/letters/062305_pachauri.pdf.

Jansen, E., Overpeck, J., Briffa, K. R., Duplessy, J. C., Joos, F., Masson-Delmotte, V., Olago, D., Otto-Bliesner, B., Peltier, W. R., Rahmstorf, S., Ramesh, R., Raynaud, D., Rind, D., Solomina, O., Villalba, R. and Zhang, D. (2007). Paleoclimate. In *Climate Change 2007: The Physical Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and H. L. Miller, eds.). Cambridge Univ. Press, Cambridge.

Mann, M. E., Bradley, R. E. and Hughes, M. K. (1998). Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* **392** 779–787.

Mann, M. E., Bradley, R. E. and Hughes, M. K. (1999). Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophysical Research Letters* **26** 759–762.

McIntyre, S. and McKitrick, R. (2003). Corrections to the Mann et al. (1998) proxy base and Northern Hemispheric average temperature series. *Energy and Environment* **14** 751–771.

McIntyre, S. and McKitrick, R. (2005a). Hockey sticks, principal components, and spurious significance. *Geophysical Research Letters* **32** L03710.

McIntyre, S. and McKitrick, R. (2005b). The M&M critique of MBH98 and Northern Hmispheric climate index: Update and implications. *Energy and Environment* **16** 69–100.

McIntyre, S. and McKitrick, R. (2005c). Reply to comment by Huybers on "Hockey sticks, principal components, and spurious significance." *Geophysical Research Letters* **32** L20713.

McShane, B. B. and Wyner, A. J. (2011). A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Ann. Appl. Statist.* **5** 5–44.

National Research Council (2006). Surface temperature reconstructions for the last 2000 years. Committee on surface temperature reconstructions for the last 2000 years.

Board on Atmospheric Sciences and Climate, Division of Earth and Life Studies. The National Academies Press.

Sir Russell, M., Boulton, G., Clarke, P., Eyton, D. and Norton, J. (2010). The independent climate change e-mails review.

Tett, S. F. B., Stott, P. A., Allen, M. R., Ingram, W. J. and Mitchell, J. F. B. (1999). Causes of twentieth-century temperature change near the Earth's surface. *Nature* **399** 569–572.

Wahl, E. R. and Amman, C. M. (2007). Robustness of Mann, Bradley, Hughes reconstruction of the Northern Hemisphere surface temperatures: Examination of criticisms based on the nature and processing of proxy climate evidence. *Climatic Change* **85** 33–69.

Wegman, E. J., Scott, D. W. and Said, Y. H. (2006). Ad Hoc Committee Report on the "Hockey Stick" Global Climate Reconstruction. Report Delivered to the Chairman of the United States House of Representatives Committee on Energy and Commerce and the Chairman of the Subcommittee on Oversight and Investigations, July 14. Available at: http://energycommerce.house.gov/108/home/07142006_Wegman_Report.pdf.

NOAA/NESDIS/NCDC/
Paleoclimatology Branch
325 Broadway Street
Boulder, Colorado 80305
USA
E-mail: Eugene.R.Wahl@noaa.gov
URL: http://www.ncdc.noaa.gov/paleo/paleo.html

National Center
for Atmospheric Research
1850 Table Mesa Drive
Boulder, Colorado 80305
USA
E-mail: ammann@ucar.edu
URL: http://www.cgd.ucar.edu/ccr/ammann/millennium/MBH-reevaluation.html